DOCUMENT RESUME

ED 110 536 UD 015 341

AUTHOR Brimer, M. A.
TITLE Evaluation Research and Action Programmes Amongst the Educationally and Socially Disadvantaged. Technical Report.
INSTITUTION Organisation for Economic Cooperation and Development, Paris (France). Centre for Educational Research and Innovation.
REPORT NO CERI/DP/71.01
PUB DATE Sep 71
NOTE 27p.

EDRS PRICE MF-$0.76 HC-$1.95 PLUS POSTAGE
DESCRIPTORS Economically Disadvantaged; Educationally Disadvantaged; *Educational Programs; Evaluation Criteria; Intervention; Models; Poverty Programs; *Program Evaluation; *Program Planning; *Research Methodology; Research Needs; Research Problems; *Social Action; Socially Disadvantaged

ABSTRACT

The stated purpose of this paper is to review some of the critical issues that arise in the planning and conduct of the evaluation of action programmes amongst the socially and educationally disadvantaged, and to suggest the means that might be adopted to overcome the inherent difficulties. The discussion first focuses on the characteristics of social and educational action programmes. Then, the functions of evaluative research are discussed. Following this, the discussion asserts that the coexistence of educational and social disadvantage in extreme forms tends to evoke programs which make either one or several interventions and assume that the interventions will make the social system work more effectively. It is advocated here, that in the planning stage, a model be constructed into which the proposals for change can be fitted and, through which, how they work and their intended effects can be seen. The value of such models, it is stated, is that they allow each proposal to be worked out in terms of its specific objectives, the agencies of the system and the instruments of action. The discussion then focuses on "the choice and formulation of objectives" and "the special status of evolving programmes". Finally, the discussion highlights considerations regarding "the design of evaluation experiments" and "the relationship between research and action".

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED110536

centre for educational research and innovation

EVALUATION RESEARCH AND ACTION PROGRAMMES AMONGST THE EDUCATIONALLY AND SOCIALLY DISADVANTAGED

technical report

2

OECD ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

ED 052526

ORGANISATION FOR ECONOMIC
CO-OPERATION AND DEVELOPMENT

Paris, 15th September, 1971

Centre for Educational Research
and Innovation

Or. Engl.

CERI/DP/71.01

EVALUATION RESEARCH AND ACTION PROGRAMMES AMONGST THE EDUCATIONALLY AND SOCIALLY DISADVANTAGED

by

M.A. Brimer

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

VD 015341

82.323

Free

PREFACE

The special difficulties which are involved in providing educationally and economically disadvantaged populations with access to the fullest use of the facilities offered by the school, and thus promote equal educational opportunity, have for some time been tackled by research and experiment in compensatory education. Parts of these problems have already been mentioned in the CERI publication "Equal Educational Opportunity"(1). A description and appraisal of selected U.S. American programmes, "Strategies of Compensation: A Review of Educational Projects for the Disadvantaged in the United States",(2) has been sponsored by CERI and will be published at a later date.

From its early beginning, CERI stressed the importance of the educational problems of disadvantaged populations and has for some years been collaborating with groups making research and experiments in pre-school education in the Netherlands, Norway, Sweden and the United Kingdom. At an early stage of this work, it appeared to be of major importance to develop evaluation practices which would lead to an increasing understanding of the value of experimental programmes in this field.

The following paper was written by Mr. M.A. Brimer, M.A., School of Education, University of Bristol, and is based on a close co-operation between the author and the members of the national groups. The author's considerable experience, gained through his work for the British E.P.A. projects, was an important advantage.

J. R. GASS
Director
Centre for Educational Research
and Innovation

---

(1) OECD, 1971.
(2) OECD, in print.

The purpose of this paper is to review some of the critical issues that arise in the planning and conduct of the evaluation of action programmes amongst the socially and educationally disadvantaged and to suggest the means that might be adopted to overcome the inherent difficulties.

Before considering the applications of evaluation to action programmes one ought first of all to look at the concepts that are applicable to the programmes themselves.

The Characteristics of Social and Educational Action Programmes

Such programmes generally accept as targets for remedial action the forms of normatively identified deprivation which themselves help to define socially disadvantaged groups. The process would be natural and proper if there were not sometimes ambiguity over the status of the variables which identify the socially disadvantaged groups. Unfortunately, it is not always certain what is the innocuous and even culturally valuable context of disadvantage and what are the variables that must be modified. The consequences of this are that action programmes tend to be multiple and that causal hypotheses are often erected on the flimsiest of concomitant variation evidence. Moreover the multiple action is concerted so that there is little hope of being able to distinguish the particular consequences of any one action or the interaction effects of the multiple variables that are being manipulated.

Potentiality for confusion is added to ambiguity in so far as the basis of both identification of such groups and the wish to institute action arises from a moral stand-point with ideological association. The concepts of equality or of equality of opportunity support the moral judgement that action is necessary and introduces a notion of normative control. Objectives are defined in terms of normative standards, the difference between sub-samples of the normative group being used to justify the need for particular forms of action related to the centrally tending targets. Discrepancies observed at one point in time are not the result of short-term effects alone but often of generations of change. The long-term changes which have resulted in a particular group of people being in a less advantageous position than others are often very diverse. Nevertheless the effect of different social and environmental influences is to

bring the disadvantaged together into one apparently homogeneous group (either because of their common location or being regarded together because of their demands upon the community in a similar form e.g. intensive poverty) and for their problems to be considered as if they arose from common causes and without regard to the historical and biographical determinants of their current common disadvantage. A further complicating factor arises when there is a geographical concentration of such groups such that the interaction of their common symptoms of itself produces phenomena which reinforces the social scientist's view of the common origin of their problems.

Since the disadvantaged are commonly recognisable in terms of the absence or lower level of certain attributes which characterise the normatively secure groups, it is a natural and rather naive step to then formulate compensatory programmes which seek to make good the deficiencies. So far as educational action is concerned it is agreed that the earlier children are treated the less likely it is that their domestic deprivation will be allowed to become a progressive handicap and it is not unnatural that the action should be modelled on what is considered to be characteristically good practice in normatively good domestic situations. Thus some of the commonest forms of action programmes are really attempts to simulate the conditions of a good home in the more limited circumstances of an institutional pre-school setting. The naivety of such an approach consists not least in the failure to perceive that the effects which are characteristic of good homes and the actions which are associated with them are the result of long-term exposure. It is not merely that the number of hours per day that a child spends in the context of a good home has its effect but also that there are regularities of context and personal interaction which extend over a long period usually unbroken. Attempts to simulate such characteristics will always be handicapped by the inability to reproduce both the intensity and the extensity of exposure.

Awareness of the limitations of institutional action and of the danger of the disadvantaged parental generation exerting stronger counter influences, thereby perpetuating itself, provokes action to modify parental behaviour. But, the parents are suspicious and resentful of the alien invasion of their family life, so the community of which they are part becomes the target for action in the hope that its more compatible agencies can be enlisted and will be more effective. Action becomes widespread, diverse and disconnected.

By contrast, to be successful, the choice of particular forms of action must be dependent upon a clear perception of what is modifiable by what means, and what of that which is to be modified is regarded as instrumental in change and what is in itself an ultimately desired change effect. It is an obvious requirement that whatever is seen to be an instrumental change should be articulated with that which is considered to be an ultimate effect. Moreover the hazard of constructing chains of instrumental change, with their multiplied scope for breakdown, argues great caution and economy in their introduction. It may be preferable to devote more preparation and resources towards producing the ultimate effect directly. For example when it is considered that children in a certain group are incapable of using their first language adequately there may be a number of forms of both instrumental and ultimate effect considered as worthy of achievement. It may be decided that one instrumental effect would be a change in the form of speech that parents use towards their children. However, awareness of the difficulties of changing the speech habits of adults may lead to a decision to replace the inadequacies of the parents by simulating the conditions of a good home in an institutional setting. For the reasons already given such compensation will be unlikely to lead to change that would produce the desired effect. It may be preferable to consider what have formerly been regarded as symptoms rather than as causal factors and to change the speech characteristics of such children by intensive selective treatment. It is at this point that one may become aware that the changes which are introduced in a child's speech by such directional effort are themselves capable of being regarded merely as instrumental change. The relationship between language and thought has been instanced as the reason for seeking to improve language. If programmes are aimed merely at the symptomatic differences between children from good and bad homes the effect may be merely to change the particular and immediate fashion of speech of these children without having any long-term effect upon the way of structuring reality that the child has at his disposal. An awareness of the distinction between learning and performance would be most important at this juncture as well as the recognition that what it is desired to produce in such children is a set of structures which are capable of transfer to different situations rather than improvement of immediate competence in particular areas of experience.

A different problem of social action programmes arises from the definition of the target population, permitting as it does the disadvantaged groups to be considered in isolation from others. Apart

from the obvious recognition that no group of people ever lives in complete isolation from others of different status and that its aspirations and its achievement motivation are never formed in isolation, there is the practical difficulty that effective forms of action might, if they were applied to non-disadvantaged groups, also lead to their improvement so that the relative position of the groups would then remain the same. Once such a possibility is recognised the moral dilemma involved in normative definition of social goals increases.

There are related problems which tax the research design, amongst these being the problems of regression effect when action is directed at those who in achievement are well below the mean of a representative sample of the population.

In choosing programmes for application amongst disadvantaged groups there is a temptation to create new programmes that have not demonstrated their worth with respect to a more general population. Thus there is a tendency for programmes for the disadvantaged to be innovatory not merely with respect to the particular kinds of disadvantage exhibited but also in the sense of applying novel methodologies, previously untested on any group, let alone the disadvantaged. Just because action tends to be multiple and to involve long term treatments it is expensive. It therefore behoves any innovator to justify to himself very closely the reason for applying a particular innovation which has not been previously demonstrated to have effect within a less restricted population.

It must be apparent from the foregoing that to derive research results from experimental action amongst the disadvantaged is difficult. Indeed, it is understandable that research should tend to be forgotten in the welter of action and that evaluation should be regarded as a means of justifying action.

Evaluative Research

The functions of evaluative research are rather closely definable and they must be distinguished in the first place from those of mere evaluation. Evaluation of itself only presupposes the assignment of value to anything and does not prescribe the need for any scientific rigour in the process of assigning value. Evaluative research on the other hand implies the use of a methodology, the statement of a set of premisees, the conformity with a set of rules and obedience to certain criteria. Bigman (1961) has listed the following main functions of evaluative research.

1. To discover whether, and how well, objectives are being fulfilled.
2. To determine the reasons for specific successes and failures.
3. To uncover the principles underlying a successful programme.
4. To direct the course of experiments with techniques for increasing effectiveness.
5. To lay the basis for further research on the reasons for the relative success of alternative techniques.
6. To redefine the means to be used for obtaining objectives, and to redefine subgoals in the light of research findings.

Implicit in such a statement of function is the recognition that action and research must be planned together and that the business of evaluative research is as much concerned with the means as with the outcomes. It further requires that context be specified so that it is possible to disentangle the implications of particular contextual characteristics in their interaction with particular aspects of the programmes. This notion of context also has implications for sampling.

If, for example, the target population is all urban children between the ages three and five, whose per capita family income falls below a given level, the context of their homes and of their cultural milieu is as important as age and income level for generalising the outcomes of the study. Unless each neighbourhood of each city is to be regarded as containing an independent target population, there must be regard for generalisable aspects of context. In other words, the target population must be stratified for whatever contextual characteristics are regarded as sufficiently important to consider separately. There will be many more contextual characteristics which are unsuitable for stratification but which will be observable and measurable within the sample. Multiple interactions of these characteristics with each other and with aspects of the programme will occur and are potentially vital in accounting for the events of the study.

When such interactions between context and programme are being considered it becomes necessary to formulate a model for the existing interaction within the contextual system and to specify the hypotheses of the effective change upon such interaction. Unless this is done it becomes impossible to test the multiplicity of possible hypotheses that might have been erected when employing a probability model.

Multiple Action Programmes and Dynamic Context Programmes

The coexistence of educational and social disadvantage in extreme form tends to evoke multiple action programmes or generalised single action programmes in which it is supposed that the effect of the action will be to increase the output of the multiple elements of the social system through its own dynamics. Such programmes are characterised by multiple objectives and they assume the interaction of agencies of instrumental change which may have a combined effect upon the realisation of objectives. Optimism about the combined effects of change or of the organic growth of the single change frequently prevents the specification of the particular interactions which are anticipated and, perhaps more important, restricts the perception of those forms of interaction which may be disadvantageous as well as advantageous to the overall achievement of goals.

The only possible way of dealing with this situation is to set up at the planning stage a model into which the proposals for change can be fitted and through which their mode of operation and their intended effects can be seen. If the model can reveal interaction between agencies such that the combined effect of changes can be estimated so much the better. Where educational disadvantage is the focal symptom under scrutiny, it is reasonable to project the programme's objectives in the educational terms which are focally important, although it may be that for other purposes or for a further prescription of another part of the model one might set up social objectives as being instrumental in educational change. Normally a comparison of the programme under test with another programme which is regarded as standard, or as differing in some controlled fashion, is considered to be a vital part of the evaluation. It follows that the model should contain sufficient flexibility to enable both sets of objectives arising from the standard and from the experimental programme to be encompassed. A convenient model which illustrates the way in which the system may be constructed at the planning stage is that derived from Carroll's (1963) model of school learning. It also involves the notion that all learners are capable of achieving mastery given sufficient time and therefore all but the criterion variable are measured in terms of time for learning. The elements of the model might be defined as follows:

1. Perseverance - readiness of the learner to invest actively in learning
   brought about by (a) level of motivation
   (b) degree of positive attitude.
   Measured through amount of available time utilised.

2. Aptitude of Learner - readiness of the learner for a particular kind of learning brought about by
   (a) level of general ability
   (b) knowledge already possessed.
   Measured through time taken to learn under optimal conditions.

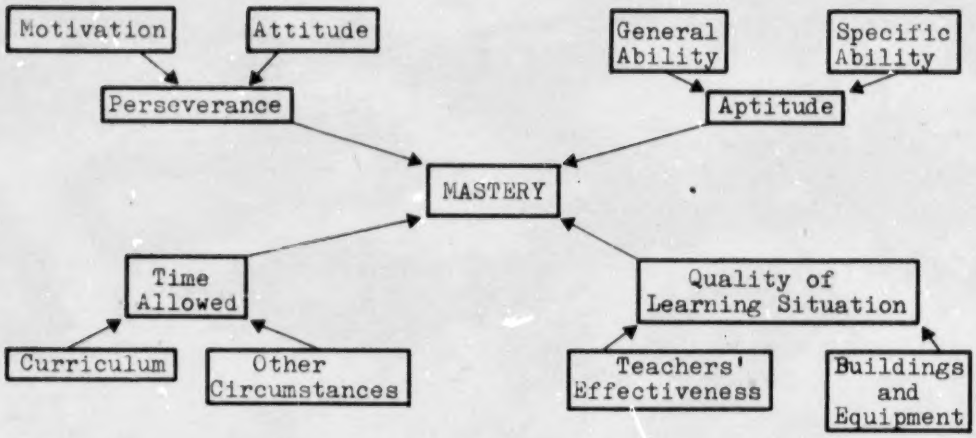
3. Time allowed for learning - created by
   (a) curriculum
   (b) the disposition of other circumstances which might limit or increase curriculum time.
   Measured through time provision.

4. Quality of the Learning Situation - dependent upon
   (a) the effectiveness of the teacher
   (b) the provision and organisation of buildings and equipment.
   Measured by time required for learning in excess of that minimally required by aptitude.

5. Skill mastery - achievement of goal objectives measured through criterion - referenced tests.

(The intention here is to escape from norm referenced tests which are incapable of specifying the extent to which the curriculum objectives have been achieved.)

The model is diagrammed as follows - lines show the links and arrows the direction of effect.



Each one of the input elements to the model is capable of further specification in terms of its own system where it falls into the centre of the system and so on to the boundary of the intended scheme.

For example, teaching effectiveness might itself become the focus of a system prescription and the elements of the model might be defined as follows:

1. Perseverance - identification with aims of project.
   Dependent on (a) Motivation
   (b) Attitude
   Measured through time devoted to project.

2. Aptitude - ability to contribute towards achievement of aims of project.
   Dependent on (a) General level of ability
   (b) Specific skills and knowledge.
   Measured through time required above that minimally possible to inculcate skills and knowledge.

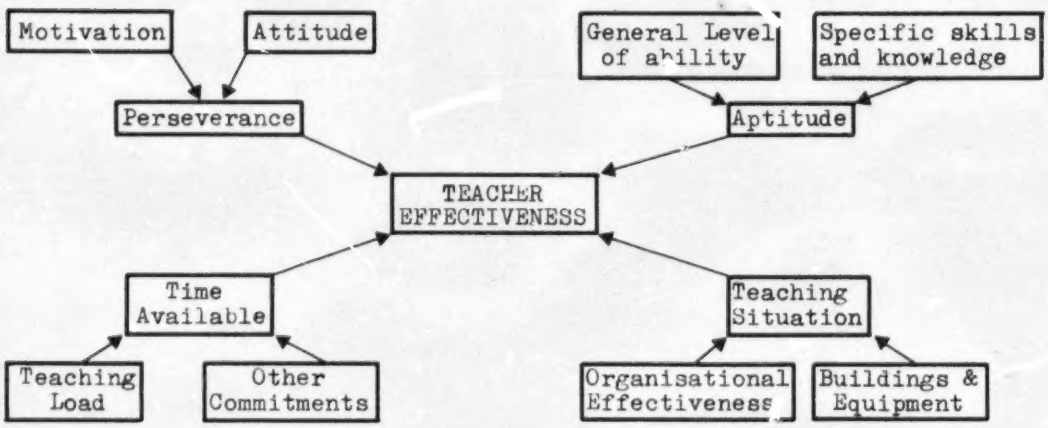
3. Time available for increasing Teacher Effectiveness.
   Dependent on (a) Teacher load
   (b) Other commitments.
   Measured through time without inescapable other commitment.

4. Teaching Situation

Dependent on (a) Organisational effectiveness

(b) Buildings and equipment.

Measured through time required above that minimally possible to inculcate skills and knowledge.



Clearly the models can be articulated further but it is often preferable to leave them at as general a level as one can in the early planning stage in order to aid coherent scrutiny.

The value of such models is that they allow each proposal to be worked out in terms of its specific objectives, the agencies of the system and the instruments of action, making explicit intentions for action and hypotheses of effect. Successively more detailed specifications are made within the model until a defined programme emerges.

The Choice and Formulation of Objectives

The introduction has already exposed the tendency of projects aimed at the disadvantaged to elevate concomitant differences between good and poor homes as arguments both for action and for the objectives to be sought. While it is obviously true that objectives can only be identified through an appeal to a value system, the values themselves do not prescribe strategies which enable objectives to be selected or their status to be recognised as part of a causal chain. The proliferation of objectives arising from an indiscriminate appeal to a value system can frustrate action, by diffusing its processes, and evaluation by substituting multiple vague hopes for

explicit expectations. In seeking to explain the inadequacies of action programmes, criticism has been strangely levelled at the limitation of the forms of action carried out. (See, for example, Tizard 1970). Thus broader value systems are drawn in, which specify objectives which have previously been conceived to be peripheral to the action and its objectives. Argument centres on the relative priority status of values and the restrictions of values chosen rather than upon the structural relationship of the sources of variation which might be manipulated. The application of a model such as that suggested in the last section will help to locate objectives and to screen them for their structural and causal status in relation to one another. Since all action programmes imply the existence of causal relationships it is necessary to appeal to a body of theory which is more general than the particular hypotheses under test.

It is highly likely that in such programmes, competing, or apparently competing, theories will emerge. In particular there may be seen to be a competition between those theories which imply a sociological view of the system and those which imply a psychological view. Reddiford G. (1969) has cogently argued the independence of sociological concepts from those of psychology. It follows from such a thesis that sociological concepts are not themselves explicable entirely in terms of psychological concepts and it is possible to entertain both within the same model on a non-competitive footing. There are rather more serious questions as to whether sociological concepts and constructs can ever be causal in character. Thus it may be necessary to specify some objectives in sociological terms but to turn to psychological constructs for the purposes of securing change which the sociologist would then recognise in terms of his own modes of testing social systems. Whatever the means chosen for reconciling the apparently competitive demand of social and behavioural scientists it is vital that the form of reconciliation should not merely be one of tolerance of forms of independent action which cannot be related.

Strategically it is always helpful to set out the categories of ultimate goal before considering intermediate goals even though action may be directed at intermediate and instrumental change. To leave ultimate goals unexamined is to risk substituting within-system efficiency for pay-off from the system. The categories of ultimate goal which might be screened are those which are considered to be the change areas for action: such categories as knowledge, skills,

attitudes, social interaction and role perception may be set up as possible groups of outcomes. Under each of these headings the major objectives are listed and then their subordinate objectives formulated until the point is reached when exemplars in an operationalised form are written. It is a considerable aid to employ a system such as the Taxonomy of Educational Objectives (Bloom, 1956), whenever the objectives form part of an orderly universe. It not infrequently happens that the objectives are not so easily specified within a hierarchic structure and this is most apparent when it is a construct which identifies the area in which change is to be sought. Constructs by their nature cannot be directly sampled and can only be monitored through indirect observations which are inferential deductions from the construct theory.

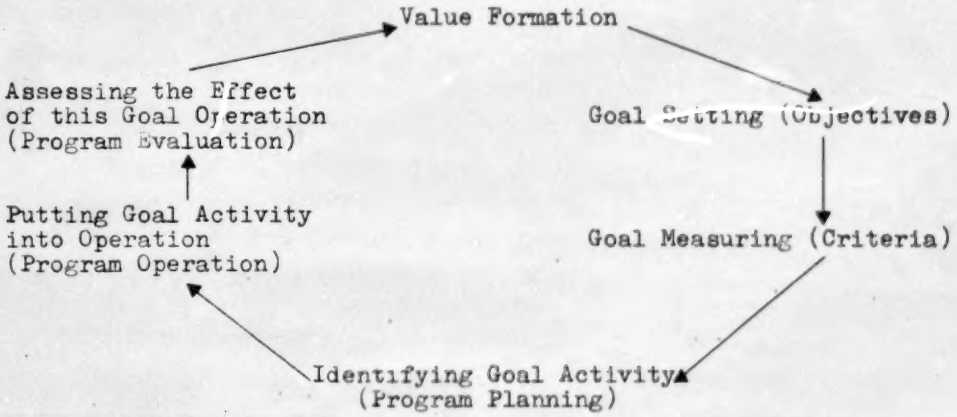
An example of such a construct would be in terms of the area of attitude referred to above. As a construct, 'attitude' refers to the coherent structure of action and reaction tendencies possessed by an individual or a group of persons related to a particular object, the structure being relatively stable and resistant to change. It is considered to be affectively rather than cognitively determined although capable of rational explication and to be associated with beliefs held which are independent of veridical evidence. Attitudes are therefore most commonly measured by inviting judgements of degrees of personal conformity with verbal statements expressing opinion about the object. To sample the universe of aspects of the object or the universe of verbal statements of opinion about aspects of the object would clearly be insufficient as a basis for testing the construct itself. Therefore scales are constructed which accord to responses to statements, scale values which are homogeneous and therefore additive. Such observation of responses is clearly indirectly related to the construct itself and individuals may vary in their response to verbal statements without thereby implying a shift in the status of their attitudes. For example, it has been a common finding in England that the attitudes of teachers towards education and children tend to become more conservative and tough-minded after a period in teaching than they were when the teachers were students. Such a shift in observed responses could equally well be explained by according to the teachers a greater tolerance of verbal statement of conservative and tough-minded opinion as not necessarily implying conservative or tough-minded action. The lesson to be learned from this is that confidence in the fluctuation of observations as implying variation in the construct under test can

only be engendered if other sources of variation in observation have been controlled.

The Special Status of Evolving Programmes

A distinction should be recognised between evaluation applied to already institutionalised programmes, newly applied but previously tested programmes and untried evolving programmes. All too often these aspects are inextricably mixed up together. While the first two types of programmes have much in common the third is very different and demands special consideration.

The characteristic of an evolving programme for the evaluator is that it requires short term feed-back of information on the basis of which the programme will itself change. It also needs to create provision so that modified parts of the programme can be reintroduced in a different setting without bias. Suchman (1967) has diagrammed a list of procedures that are necessary.



According to this model the procedure recirculates until a point is reached where the system can produce an exit, that is to say, when the assessment of the effect of the whole operation does not lead to a redefinition of values and programme objectives but rather points the way to an extension of the objectives in an 'end-on' form. One of the main differences in the evaluation of an evolving programme is therefore that the assessment is related to decision-making about values and objectives as well as to programme efficiency. The requirement of short-term feed-back of evaluation consequences imposes restrictions on the type of evaluation that can be carried out. In general this means that forms of evaluation must be sought which

maximise validity but which are more tolerant of variation in reliability. Programme planners must work in a manner which is not unlike that of curriculum builders in that they must seek to anticipate decision points concerning alternative ways of conducting steps in the programme. Evaluation would then be closely geared to a decision between competing modes of conduct. Glaser (1967) has outlined the steps which are likely to be necessary in a curriculum building exercise pursuing this model. Briefly, the model requires the hierarchic specification of the objectives in which intermediate objectives are expressed in terms of their dependence upon earlier, and in terms of their contribution to later, objectives; the specification both of the achievement and the learning characteristics of the learner; the selection of the most promising learning situations and procedures for competitive experiment and evaluation which is aimed at testing the goodness of the objectives, the relative success of the competing learning procedures and the goodness of fit of the procedures to the individual learner's characteristics.

In evolving a programme the commitment is essentially long-term but the form of experiment is not itself long-term. Apart from the ethical and institutional problems of pursuing novel procedures with the same group of subjects over a long time there is the question of the uneconomic use of time in the development of the programme if every stage must be dependent upon the prior administration of an earlier stage. If objectives are appropriately operationalised and criteria exist whereby individuals can be pre-tested to determine their degree of satisfaction of these objectives it becomes possible to carry through a series of simultaneous experiments in which various steps of the programme are independently tried and only at a later stage fitted together. The benefit of this method is not merely economy of time but also that it prevents the pursuit of inadequate instrumental objectives if later occurring objectives have been found to be inadequate. A further benefit according to this model is that it prevents the specification of programmes which have some spurious or accidental connection with the achievement of objectives. It becomes possible within such an experimentally developed system to specify a causal relationship between the elements of the programme and the objectives achieved. The benefit is not merely in terms of the validity of the programme but also in terms of the degree of conviction that is possible amongst those who are responsible for mounting it.

The Design of Evaluation Experiments

It would be inappropriate to attempt a complete study of experimental design in so far as it applies to evaluation research in such a paper. There has indeed been no dearth of publications which set out the problems and processes clearly. Perhaps one of the reasons why the wealth of information available has been heeded so rarely is that discussion of experimental design is linked with statistical formulae and mathematical probability which are not comprehensible to all those who mount evaluation of action programmes. It might be asked why those who write such books have not themselves taken more trouble to make their influence felt directly in the construction of evaluative research. The reason is probably that the expert in research design so often finds himself intolerant of the apparent untidiness of action programmes and his partners in the exercise despair of matching the elegance that the research design expert requires. Such mutual frustration is even less conducive to a compatible partnership when evaluation is seen as a distant requirement of a report subsequent to the study and when the short term feed-back to the action director is heavily biased in favour of functional viability of the action.

Evaluative research designs can be chosen on the basis of a compromise between the demands of exhaustive testing of hypotheses and practical feasibility. The nature of the compromise is indicated by the degree of exclusion of alternative hypotheses that are minimally required in order to carry public conviction to be a point of extending the action to larger samples of the population. It follows that the degree of sophistication of the evaluative research must be subject to discussion by the planning team and is not solely within the competence of the research worker. It is the duty of the research worker to reveal clearly to the team the consequences of failing to adopt a more sophisticated design.

It should be recognised at the beginning of the planning of evaluative research that no experiment is ever completely free from error. The function of experimental design is to enable the error to be known and to be kept within the limits that the study will tolerate. The main categories of error that design seeks to control are:

1. Variation attributable to subjects.
2. Variation improperly associated with treatment.
3. Variation attributable to group-specific reaction to treatment.

Error of type 1 arises when the subjects who are being subjected to the treatment are themselves a biased sample of the target population. Errors of type 2 arise when the treatment under test is associated with biased conditions which may be a function:

- of the treatment itself, e.g. novelty;
- of the physical circumstances, e.g. the classroom;
- of those who are instrumental in presenting the treatment, e.g. the teacher;
- of the social conditions in which the experiment takes place, e.g. the climate of the school;
- of the groups to which the treatment under test is applied, e.g. the state of restlessness of the groups.

The third type of error arises when the treatment under test reacts with some prior biased condition of the groups to whom the treatment is applied, e.g. when the curriculum previously followed is particularly conducive to the treatment under test.

The mode of controlling for these errors is of two basic kinds. The object of both is to eliminate the effect of the error from the evaluation of the experimental trial. The first control mode is concerned with the use of prior knowledge about the possible sources of variation such that they can be eliminated or counterbalanced. The second mode involves the principles of randomisation, the effect of which is to eliminate directional bias in the allocation of subjects to groups or of treatments to groups. Although randomisation avoids directional bias it nevertheless admits the possibility of bias arising by chance. The virtue of randomisation is that the variation which can arise from chance can be accountable in the tests of significance.

An excellent account of research design as it applies to evaluation studies is given in "Evaluative Research" by E.A. Suchman, Russell Sage Foundation, New York, 1967, Chapter 6. It is unnecessary to cover the same ground again. One of the purposes of this paper is to suggest those designs which are likely to maximise outcome from evaluation under the characteristic conditions which surround compensatory programmes. Three sets of conditions are posited which in themselves are stages in the strategy of programme preparation and testing.

The first condition arises when a particular region or class of deprivation is being considered for programme preparation. The main objective of this stage is to determine what forms of deprivation exist and with what conditions it is linked, and further, what hypotheses of causal relationship and of interaction are tenable. As such, it is a survey in which a representative sample of the target population is studied and the resulting observations are subjected to multivariate analyses. The techniques of analysis to be applied are such standard forms as those described by Hope,K., "Methods of Multivariate Analysis", University of London Press, 1968. The particular techniques that apply to the examination of causal relationships in non-experimental data are relatively more recent in development (Blalock, H.M. 1963 and 1964; Duncan, O.D. 1966; Keesling, J.W. 1968; Tukey, J.W. 1954; Wold, H. 1956 and 1967; Wright, S. 1934 and 1960; Yee, A.H. and Gage, N.L. 1968). The planning of the survey requires the same care in specifying the conditions which are to be investigated as an experimental study would do. The outcome of the analyses should be the hypotheses to be tested and the indicators or variables to be used.

Results of such studies would naturally lead to the adoption of certain hypotheses of change agencies and instructional routes which could then form the second stage in which experiments would set out to test the hypotheses. The conditions associated with this stage are likely to be characterised by tentativeness of hypotheses to be tested and perhaps of a number of alternative hypotheses to be tested. It would be futile under these conditions to embark on a long-term gamble with the lives of so many children and so much effort. It is preferable to focus strategy on decision making about those routes and those agencies which are most likely to be profitable in long-term study. It follows that at this stage the experimental work should be short-term, should employ limited target populations, employ small samples and should again seek to maximise outcome while minimising error. The host of disturbing effects in educational and social settings is such that the dice should not be unnecessarily loaded against the treatment under test. In many cases the attempt to maximise the outcome of interaction of treatments with individual differences frustrates experiment by imposing impossible demands upon the constitution of groups. Matching techniques, whether by individuals or by levels, are not only increasingly difficult to implement as the number of determining variables and their levels increases but tends to diminish the size of the samples under test

irrelevantly and to restrict the variance of any one variable and to make the experiment dependent upon the premisses of the foreseen covariance when other unknown covariance may arise which is a function partly of the restrictive effect of representing each variable at each level of every other.

The suggestion is that a number of simultaneous short-term experiments should replace the attempt to mount the long-term experiment immediately. An example taken from the present author's own work (Brimer, M.A. 1967) may serve to illustrate the mode of tackling the problem. The objective of the experiment was to discover the relative effectiveness of five orthographies in learning the code of the traditional English orthography. The design chosen aimed at the control of differences in initial ability of children, differences between schools and teachers, the effect of change and novelty and of the method of instruction. The children were selected at random from the appropriate age group within each of five schools and were allocated at random to one of three groups of eight to be taught through a particular orthography. The children were also pre-tested on a measure of listening comprehension so that covariance methods could ultimately be used to increase the precision of the randomised design. Orthographies were allocated to schools at random under the condition that each school should have three different orthographies and that each orthography should be followed in three different schools. By this means it was hoped to control the initial ability of the children in association with orthographies and to eliminate school bias. In each school one teacher was responsible for instruction with all three orthographies. The effect of this was to control teacher bias to the same extent that school bias was controlled. A further control on teacher effectiveness and an equalising of the effect of novelty was sought by introducing a form of programmed instruction involving simple teaching machines. The method of instruction also permitted a continuous record to be kept of the performance of each of the groups at each stage of the learning. Thus it was ultimately possible to compare not only the final achievement on each orthography but also to compare the rates of change of performance across orthographies. The actual instruction period was no more than fourteen days, yet the data derived were such that adequate discrimination between the orthographies could be measured.

By strategically mounting a number of simultaneous studies of no greater duration than this it would be possible to solve a number of crucial problems in relatively isolated conditions before fitting

the contributory processes together into an articulated programme. Such a process would be much less dangerous if it were subsequent to the survey described earlier in which causal paths had already been tested on a non-experimental basis.

The completion of this second stage would then lead naturally to the third stage of mounting the long-term study which would be set in real circumstances. In contrast to the second stage, the target population would be fixed bearing in mind the desirability of being able to generalise beyond the confines of limited local circumstances and the samples would be drawn in such a way that they represented the target population in its naturalistic environment. In most cases this would mean the acceptance into the study of intact groups without the capability of being able to randomise subjects in relation to groups. Covariance methods would be used to control subject-linked, group variation and treatments would be assigned randomly to the intact groups. Covariance methods would permit greater precision to be achieved in the experiment through control of residual variation between randomised treatment blocks. Evaluation would be carried out on group means and would discount individual variation within groups. If the treatment were such that it were to be adjusted to intermediate characteristics of the learner then the newly constituted groups would become the unit of analysis rather than the individual learner. The constitution of such groups for treatment and analysis purposes does not necessarily imply that they are taught in the same physical location but rather that a common attribute of treatment is being evaluated for its consequent change effect upon criterion-referenced measures.

The Relationship between Research and Action

It is accepted as almost inevitable in the evaluation of action programmes that research workers and action directors should be at variance with one another. In part this has seemed to arise from the different objectives which they each pursue, in part through the degree of impartiality towards the action that they represent and in part through the degree of involvement in the action that they each concede. There is a further variation which is attributable to their degree of involvement in context. The action workers are inevitably involved in the immediate day to day context of the experimental situation whereas the research worker is to a large extent insulated from it. The difference also arises through the kind of competence that each of them has to offer. The success with which

action workers pursue the promotion of their course of action depends upon their success in persuasion and their success in involving others in the pursuit of the desirable outcome. They tend to be social scientists who are sympathetic to the client and who are enthused by the form of action they are promoting and who regard the intervention of measurement or evaluative procedures as unsympathetic and even as unnecessary. Such people are, on the whole, not inclined towards quantifying or measuring whereas the research worker is essentially bound up with the conformity of action with the prescribed design, with the maintenance of a rigorous schedule of assessment and ultimately with the desire to demonstrate the efficiency of the procedures and their impartiality rather than to demonstrate the success of the form of action.

These differences are often exacerbated by the relationship within the project of the action workers and research workers in an organisational sense. Not infrequently the research workers are called-in in a consultant capacity and may not belong to the organisation which is promoting the action. Not unnaturally this tends towards creating friction which is nevertheless still evident when it is a single institution which is promoting action and monitoring the consequences. The differences between the personnel within a single institution may be even more institutionalised and extreme than those between institutions. Prior agreement on the degree of constraint which the evaluative research can be allowed to exercise is no guarantee of harmonious relationship. Such agreement is usually gained at the point at which ambitions are high and difficulties are minimised. The situation changes markedly as soon as the fieldwork begins to take place and the action workers find themselves in certain difficulties in promoting action when they are seeking the co-operation of the host of agencies which relate to an institution and the co-operation of the members of the institution itself. It is often easier for action workers to accede to adaptations of the form of action to institutional stresses than it is for research workers to agree to similar changes. No one can pretend that any means exists to eliminate all such difficulties. One mode which will reduce such sources of friction requires planning in such detail that normally it would be comparable with the organisation of a moon-landing. The construction of detailed flow diagrams, personal action charts and fail-safe procedures would then be supported by pilot studies and simulation exercises. Very rarely is sufficient time spent in planning for such procedures to be accomplished. Frequently the source of difficulty lies in the way of

handling the project where it is incumbent on those who are going to carry it through to begin it as quickly as possible in order to be able to work within the budget permitted. There is nevertheless another source of premature action in the impatience of those who are connected with the project to get ahead with it rather than to plan through all the steps that are necessary.

The evaluative research worker can never be in an entirely happy position; F.G. Caro (1969) in a recent paper claims that "the researcher walks on a tight rope between an affiliation with power groups and the freedom to hear all sides without becoming involved in internal conflicts." His very attempt at impartiality may pose a further problem. C. Argyris (1960) has argued that a research worker who is more directly involved in a programme is more effective than one who represses his humanitarianism in favour of the demand for research. He even suggests that his attempt at impartiality leads to conflict with the action worker and that this conflict in its turn tends to render his observations untrustworthy. Ambiguities of status and role, the conflicts between the demands of the research worker for records and the tedium of record keeping and gathering, all impose constraints upon the research worker for which he must be prepared not merely in the sense of being willing to accept such effects but also in terms of planned procedures to deal with the consequences.

Perhaps it is in the budgetary field that the research worker and action worker may find themselves most at variance. To the action worker the expense of research may seem to be disproportionate to the amount of effect it produces. Research information is often perceived by the action worker as being inert and it is likely that those who are administering projects will look more sympathetically at demands for further budgetary provision from those who are able to justify their requests on humanitarian grounds rather than from those who require the money for evaluative purposes. It frequently happens that evaluation budgets are cut in order to supplement action.

Conclusion

It is not possible in a paper of these dimensions to emulate a handbook on evaluative research. The purpose of this paper has been to draw attention to the most profitable strategies and to refer to texts which are by their nature rather more detailed than this paper could hope to be. Reference has already been made to "Evaluative Research" by E.A. Suchman and another general text is in press which should complement it. The anticipated work is "The Evaluation of Instruction: issues and problems", edited by M.C. Wittrock and D.E. Wiley Holt Rinehart, and Winston, 1970.

REFERENCES

Argyris, C., Understanding Organizational Behavior, Dorsey, Homewood, Illinois, 1960; Lippitt, Watson and Westley, op. cit.; Rodman and Kolodny, op. cit., p.97.

Bigman, S.K., "Evaluating the Effectiveness of Religious Programs," Review of Religious Research, vol. 2, Winter, 1961, p.99.

Blalock, H.M. "Making Causal Inferences for Unmeasured Variables from Correlations among Indicators" American Journal of Sociology, LXIX, July, 1963, 53-62.

Blalock, H.M., Jr. Causal inferences in non-experimental research. Chapel Hill: University of North Carolina Press, 1964.

Bloom, B.S., (Ed.) Taxonomy of Educational Objectives: Handbook I: Cognitive Domain. New York: McKay 1956.

Brimer, M.A., "An Experimental Evaluation of Coded Scripts in Initial Reading", New Research in Education, Vol.1, 1967.

Caro, F.G., Approaches to Evaluative Research: A Review, Publication No.111 of the Institute of Behavioral Science, Vol.28, No.2, Summer 1969, 87.

Carroll, J.B., 'A Model of School Learning', Teachers College Record 64, 8 May, 1963.

Duncan, O.D., "Path Analysis: Sociological Examples" American Journal of Sociology, LXXII, July, 1966, 1-16.

Glaser, R., "Evaluation of Instruction and Changing Educational Models", University of Pittsburgh - Learning R & D Center, Reprint 46, 1967.

Hope, K., "Methods of Multi-variate Analysis", University of London Press, 1968.

Keesling, J.W., Path Analysis: A Maximum Likelihood Solution, Abstract, 1968.

Reddiford, G., Sociology, Psychology and the Study of Education, Proceedings of the Philosophy of Education Society of Great Britain, 1969, Vol.3.

Suchman, E.A., Russell Sage Foundation, New York, "Evaluative Research", 1967.

Suchman, E.A., "The Evaluation of Instruction: issues and problems". Edited by M.C. Wittrock and D.E. Wiley Holt Rinehart and Winston, 1970.

Tizard, J., "New Trends in Developmental Psychology", The British Journal of Educational Psychology, February, 1970, Vol.40, Part 1.

Tukey, G.W., "Causation, Regression and Path Analysis" in: Kempthorne, et al. (eds.), Statistics and Mathematics in Biology, Ames, Iowa, State College Press, 1954. Chapter iii.

Wold, H.O., Causal inference from observational data: A review of ends and means. Journal of the Royal Statistical Society (London). Series A. (General) 1956, 119, 28-50.

Wold, H.O., "Nonexperimental statistical analysis from the general point of view of scientific method". Invited paper at the International Statistical Institute, Sydney Session, August, 1967. (b)

Wright, S., "The Method of Path Coefficients", Annals of Mathematical Statistics, V, September, 1934, 161-215.

Wright, S., "Path coefficients and path regressions: Alternative or Complementary Concepts?" Biometrics, XVI, June 1960 (a) 189-202.

Yee, A.H., & Gage, N.L., Techniques for estimating the source and direction of causal influence in panel data. Psychological Bulletin, 1968, 70, 115-26.